2007 CCRTS

Title of Paper: A Probabilistic Computational Model for Identifying Organizational Structures from Uncertain Message Data

Student Paper Submission
(Suggested Track: Modeling and Simulation)

Feili Yu[*1]
E-mail: yu02001@engr.uconn.edu

Georgiy M. Levchuk
Aptima Inc.,
12 Gill Street, Suite 1400
Woburn, MA 01801
Phone: 781-935-3966x267
Fax: 781-935-4385
e-mail: georgiy@aptima.com

Krishna R. Pattipati[2]
University of Connecticut,
Dept. Of Electrical and Computer Engineering
371 Fairfield Road, Unit 1157
Storrs, CT 06269-1157
Fax: 860-486-5585
Phone: 860-486-2890
E-mail: krishna@engr.uconn.edu

This work is supported by the Office of Naval Research under Contract #N00014-00-1-0101 and #N00014-06-1-080

1. Electrical and Computer Engineering Department, University of Connecticut, Storrs, CT 06269-2155, USA.
2. Correspondence: krishna@engr.uconn.edu
*. Ph.D. Student

A Probabilistic Computational Model for Identifying Organizational Structures from Uncertain Message Data (Abstract)

Feili Yu, Georgiy Levchuck, Krishna Pattipati and Fang Tu

I. Introduction

Analysis of the structures of organizations, ranging from the more structured command systems of a conventional military to the decentralized and elusive insurgent and terrorist groups, suggests that strong relationships exist between the structure, resources, and objectives of those organizations and the resulting actions [1]. The organizational members act in their assigned missions by accomplishing tasks and these activities may leave detectable clues or events in the information space. The dynamic evolution of these events creates patterns of organizational activities, which may be related, linked, and tracked over time [3], [2]. More significantly, the patterns can be used to discover the underlying organizational structure. The observed data, however, is very sparse, which creates a challenge to connect relatively few enabling events embedded within massive amounts of data flowing into the government's intelligence and counter-terrorism agencies [4]. The information technology plays an indispensable role in supporting intelligence and knowledge discovery through collection, processing, analyzing, and utilizing these organizations' data [5]. We mean by organization discovery as the ability to recognize the command, control, communication, and task structures of the organization. However, the challenge is that most of the time we cannot observe the elements of the structures of the organization. Instead, we can obtain uncertain transaction data involving the actions and activities of the organizational members. The specific actions depend on the structure of enemy command and control organization, which in turn depend on the goals of that organization. Current approaches to detect and analyze organizational network structures, such as social network analysis (SNA), link discovery (LD) and relational data mining (RDM), and graphical inference, are inadequate to address the issue of dynamic and evolutionary nature of organizational structures. Technologies from mathematics and artificial intelligence are needed to enhance the capability of organizational structure identification.

II. Proposed modeling methodology on organizational structure identification

In this paper, we focus on identifying the mapping between hypothesized nodes of an enemy command organization and tracked individuals, their activities and their resources (used to perform those activities). Thus, our graphs are dynamically evolving, partially observed, and multi-attributed. We compute the likelihood function quantifying the belief that the observed data has been generated by a particular organization. The hypothesized organizations are predefined in the knowledge library according to available intelligence regarding similar enemy organizations, well-known structural forms from organizational theories, as well as specific existing structures that analysts propose. Thus, our framework is based on hypothesis testing principles.

We employ a Hidden Markov Random Field ($HMRF$) model and a graph matching approach to find the degree of match between hypothesized organizational structures and the observed data. A hypothesized organization and the observed data are modeled as a model graph and a data graph, respectively. The graphs are the Markov Random

Fields, where each node is correlated (related) with its neighboring nodes via links. Graph matching is performed to find a score between each pair of model and data graphs. For H hypotheses, one needs to solve H graph matching problems. The hypothesis with the highest score (rank) would be the most likely organization that could have generated the observed data.

Different from SNA analysis methods that are mostly based on qualitative and empirical studies, we quantitatively model organizational structures by employing graphical models (which will be elaborated in the following subsections) that have been extensively studied in recent years. We address the three issues identified above, which have not been addressed by link discovery and relational data mining methods, by (i) constructing hypothesized organizational structures; (ii) providing an energy function that serves as a scoring function; and (iii) developing a stochastic search algorithm to find the best hypothesized candidate structure. Therefore, our methodology has the capability to discover hidden organizational structures from real data sets.

III. Possible Solution Approaches and Expected Results

Given the Hidden Markov Random Field ($HMRF$) model for graph matching, the optimal solution is obtained by computing a mapping vector for the nodes in data graph, such that the posterior energy is minimized. Since the number of nodes in both model graph and data graph can be large (e.g., several hundred), exhaustive search over the solution space is impractical for real-world problems.

In our work, the data and model graphs have multiple attributes associated with each node and edge. In addition, the energy function, although quadratic, has dependent variables with polynomial structure. Consequently, we used classical stochastic methods finding solutions that minimize the energy function. In the methods, randomness plays a crucial role in search and learning. The general approach is to bias the search toward regions where we expect the solution to be, and allow randomness to help find good solutions and escape from local minima. There are two general classes of such methods: Simulated Annealing (SA) and Evolutionary Algorithms (EA). The former is based on concepts and techniques from physics, and the latter is inspired by the evolutionary concepts from biology. Because the SA has many successes in pattern recognition, we employ it as the solution method for our problem.

References

[1] G. Levchuk, C. Chopra , K. Pattipati, "NetSTAR: Methodology to Identify Enemy Network Structure, Tasks, Activities, and Roles", *10TH International Command and Control research and technology Symposium*, McLean, Virginia, June 13-16, 2005.

[2] H. Tu, J. Allanach, S. Singh, P. Willett and K. Pattipati, "Information Integration via Hierarchical and Hybrid Bayesian Networks", *IEEE Transactions on System, Man and Cybernetics, Part A: Systems and Humans, special issue on Advances in Heterogeneous and Complex System Integration,* vol.1, no.1, page 19-34, January 2006.

[3] K. Pattipati, P. Willett, J. Allanach, H. Tu, and S. Singh, "Hidden Markov Models and Bayesian Networks for Counter-Terrorism", in *R. Popp, J. Yen (ed). Emergent Information Technologies and Enabling Policies for Counter-Terrorism, Wiley-IEEE Press,* June 2006.

[4] R. Popp, et al., "Countering Terrorism through Information Technology", *Communications of the ACM,* Vol. 47, No. 3, pp. 36-43., 2004.

[5] H. Chen and J. Xu, "Intelligence and Security Informatics", *Annual Review of Information Science and Technology,* vol. 40, pp. 229-289, 2006.